International Journal of Mathematics and Computer Applications Research (IJMCAR)
ISSN (P): 2249–6955; ISSN (E): 2249–8060
Vol. 12, Issue 1, Jun 2022, 33–42

STUDY ON LAGRANGIAN METHODS

DR. UPDESH KUMAR[1], PRASHANT CHAUHAN[2] & PARAS BHATNAGAR[3]

[1]*K.G.K. (P.G.) College Moradabad, India*

[2]*S. L. Education Institute, Moradabad, India*

[3]*G. L. Bajaj Institute of Technology and Management Greater, Noida, India*

ABSTRACT

In this paper we consider the methods for solving a general constrained optimization problem. We discussed penalty methods for the problems with equality constraints and barrier methods for the problems with inequality constraints these are easy to combine. We consider the related, better and more complex class of augmented Lagrangian methods.

KEYWORDS: Optimization Technique, Constrained, Unconstrained Problems, Ill-Conditioning &Convergence of Series

Received: Feb 17, 2022; **Accepted**: Mar 07, 2022; **Published**: Apr 18, 2022; **Paper Id.**: IJMCARJUN20224

Original Article

INTRODUCTION

A general constrained optimization problem can be given as

$$\min_{x\in\Omega} f(x)$$
$$\Omega=\left\{x\in R^n \,\middle|\, c_i(x)=0, i\in\varepsilon, c_i(x)\geq 0, i\in I\right\}$$

Newton's method, which is much less accepted by ill-conditioning provided the linear system at each of iteration is solved directly, is accepted by having its domain of attraction shrink, hence the importance of the continuation technique.

1 Penalty Method

In this section we consider the problem

$$\min\ f(x) \tag{1a}$$

$$\text{s.t. } c_i(x)=0, i=1,2,\ldots,m \tag{1b}$$

Define

$$A^T=\left[\nabla c_1(x),\ldots\nabla c_m(x)\right] \tag{1c}$$

and assume that A has full row rank $m\leq n$.

Recall that the Lagrangian is defined by

$$L(x,\lambda)=f(x)-\sum_{i=1}^{m}\lambda_i c_i(x) \tag{1d}$$

and that the KKT conditions require, in addition to (10.1b), that at (x^*, λ^*),

$$L_x = \nabla f(x) - \sum_{i=1}^{m} \lambda_i \nabla c_i(x) = 0.$$

In the *quadratic penalty method* we solve a sequence of unconstrained minimization problems of the form

$$\min \phi(x;\mu) = f(x) + \frac{1}{2\mu} \sum_{i=1}^{m} c_i^2(x) \tag{2}$$

for a sequence of values $\mu = \mu_k \downarrow 0.$ We can use, for instance, the solution $x^*(\mu_{k-1})$ of the (k -1)st unconstrained problem as an initial guess for the unconstrained problem (2) with $\mu = \mu_k$. This is a simple continuation technique.

It is hard to imagine anything simpler to intuit. Unfortunately, however, the problem (2) becomes ill-conditioned as μ gets small. Both BFGS and CG methods become severely accepted by this. An algorithmic framework for such a method reads: Given μ0 > 0, and for $k = 0,1,2,\ldots$

Starting with x_k solve (10.2) for $\mu = \mu_k$, terminating when

$$\left\|\nabla \phi(x;\mu_k)\right\| \le \tau_k$$

Where $\tau_k \downarrow 0.$ Call the result x_k^*.

If final convergence test holds (e.g. $\tau_k \le tol$) exit

Else

Choose $\mu_{k+1} \in (0, \mu_k)$

Choose a new starting point x_{k+1}, e.g. $x_{k+1} = x_k^*$.

End

The choice of how to decrease μ can depend on how difficult it has been to solve the previous subproblem, e.g.,

$$\mu_{k+1} = \begin{cases} 0.7\mu_k & \text{if } (x;\mu_k) \text{ was hard} \\ 0.1\mu_k & \text{if } (x;\mu_k) \text{ was easy} \end{cases}$$

When comparing the gradients of the unconstrained objective function $\phi(x,\mu)$ of (2) and the Lagrangian $L(x,\lambda)$ of (1d) it appears that $-\dfrac{c_i}{\mu}$

Has replaced λ_i. Indeed, it can be shown that if $\tau_k \downarrow 0$ then $x_k^* \to x^*$

and

$$-\frac{c_i(x_k^*)}{\mu_k} \to \lambda_i^*, \quad i = 1, 2, \ldots m. \tag{3}$$

Example 1

Let us consider the minimization of the objective functions under the constraint

$x^T x = 1.$

For each value of penalty parameter μ encountered, we define the function ϕ of (2) with the line search option and tolerance $\tau_k = tol = 1.e-6,$ to solve the sub problem of minimizing ϕ. If more than 9 iterations are needed then the update is $\mu \leftarrow 0.7\mu,$ otherwise it is $\mu \leftarrow 0.1\mu$.

For the quadratic 4-variable objective function we start with the unconstrained minimizer, $x_0 = (1, 0, -1, -2)^T$ and obtain convergence after a total of 44 damped Newton iterations to

$$x^* \approx (-0.02477, 0.31073, -0.78876, 0.52980)^T.$$

The objective value increases from the unconstrained (and infeasible) $f(x_0) \approx -167.28$ to $f(x^*) \approx -133.56.$

The resulting penalty parameter sequence was

μ = 1,.1,.01,...,1.e-8,

i.e., all sub problems encountered were deemed "easy", even though for one sub problem a damped Newton step (i.e. step size < 1) was needed. The final approximation for the Lagrange multiplier is

$$(1 - x^T x)/10^{-8} \approx -40.94.$$

For the non-quadratic objective function,

$$f(x) = [[1.5 - x_1(1 - x_2)]^2 + \left[2.25 - x_1(1 - x_2^2)\right]^2 + \left[2.625 - x_1(1 - x_2^3)\right]^2,$$

We start with $x_0 = \frac{\sqrt{2}}{2}(1,1)^T$, which satisfies the constraint. Convergence to x. $x^* \approx (0.99700, -0.07744)^T$ is reached after a total of 39 iterations. The objective value is $f(x^*) \approx 4.42,$ up from 0 for the unconstrained minimum. The penalty parameter sequence was

μ = 1,.1,.01,.001,7.e-4,7.e-5,...,7.e-8,

But the path to convergence was more tortuous than these numbers indicate, as solution of the 3rd and 4th sub problems failed. The final approximation for the Lagrange multiplier is

$$(1 - x^T x)/(7 \times 10^{-8}) \approx -3.35.$$

To understand the nature of the ill-conditioning better, note that the Hessian of ϕ of (2) is

$$\nabla^2\phi(x;\mu_k) = \nabla^2 f(x) + \frac{1}{\mu_k} A^T(x)A(x) + \frac{1}{\mu_k}\sum_{i=1}^{m} c_i(x)\nabla^2 c_i(x)$$

$$\approx \nabla^2 L + \frac{1}{\mu_k} A^T A.$$

The matrix $\frac{1}{\mu_k} A^T A$ has *n-m* zero eigen values as well as *m* eigen values with size $O(\mu_k^{-1})$. So, we have an unholy mixture of very large and zero eigenvalues. This could give trouble even for Newton's method.

Fortunately, for Newton's iteration, to find the next direction *p* we can write the linear system in augmented form (verify!),

$$\begin{pmatrix} \nabla^2 f(x) + \sum_{i=1}^{m} \frac{c_i(x)}{\mu_k}\nabla^2 c_i(x) & A^T(x) \\ A(x) & -\mu_k I \end{pmatrix} \begin{pmatrix} p \\ \xi \end{pmatrix} = \begin{pmatrix} -\nabla_x \phi(x;\mu_k) \\ 0 \end{pmatrix}.$$

This matrix tends towards the KKT matrix and all is well in the limit. Finally, we note for later purposes that instead of the quadratic penalty function (2) the function

$$\phi_1(x;\mu) = f(x) + \frac{1}{\mu}\sum_{i=1}^{m} |c_i(x)| \quad (4)$$

Could be considered. This is an exact penalty function: for su. ciently small μ > 0 one minimization yields the optimal solution.

Unfortunately, that one unconstrained minimization problem turns out in general practice to be harder to solve than applying the continuation method presented before with the quadratic penalty function (2). But the function (4) also has other uses, namely, as a merit function. Thus, (10.4) can be used to assess the quality of iterates obtained by some other method for constrained optimization.

2 Barrier Method

We now consider only inequality constraints,

min $f(x)$ (5a)

s.t. $c_i(x) \geq 0, i = 1,2,...,m$ ci(x) (5b)

We may or may not have m > n, but we use the same notation A(x) as in (10.1c), dropping the requirement of a full row rank. In the log barrier method we solve a sequence of unconstrained minimization problems of the form

$$\min \psi(x;\mu) = f(x) - \mu\sum_{i=1}^{m} \log c_i(x) \quad (6)$$

for a sequence of values $\mu = \mu_k \downarrow 0.$.

Starting with a feasible x_0 in the interior of Ω we always stay strictly in the interior of Ω, so this is an *interior point method*. This feasibility is a valuable property if we stop before reaching optimum.

Example 2

Here is a very simple problem:

$$\min_x x,\ x \geq 0.$$

By (6),

$$\psi(x;\mu) = x - \mu \log x.$$

Setting $\mu = \mu_k$ we consider $0 = \frac{d\psi}{dx} = 1 - \mu / x,$ from which we get $x_k = \mu_k \to 0 = x^*.$

Clearly, as $\mu_k \downarrow 0$ we expect numerical trouble, just like for the penalty

Method in Section 10.1. The algorithmic framework is also the same as for the penalty method, with ψ replacing ϕ there.

Note that

$$\nabla_x \psi(x,\mu) = \nabla f(x) - \sum_{i=1}^{m} \frac{\mu}{c_i(x)} \nabla c_i(x).$$

Comparing this to the gradient of the Lagrangian (10.1d) we expect that, as $\mu_k \downarrow 0,$

$$\frac{\mu_k}{c_i(x_k^*)} \to \lambda_i^*, \qquad i = 1, 2, ..., m.$$

For the strictly inactive constraints $(c_i > 0)$ we get $\lambda_i^* = 0$ in (7), as we should.

It is easy to see that, sufficiently close to the optimal solution,

$$\nabla^2 \psi(x;\mu_k) \approx \nabla^2 L(x^*,\lambda^*) + \sum_{i=1}^{m} \frac{(\lambda_i^*)^2}{\mu_k} \nabla c_i(x) \nabla c_i(x)^T$$

$$= \nabla^2 L(x^*,\lambda^*) + \sum_{i \in A(x^*)} \frac{(\lambda_i^*)^2}{\mu_k} \nabla c_i(x) \nabla c_i(x)^T. \tag{7}$$

This expresses ill-conditioning exactly as in the penalty case (unless there are no active constraints at all).

Let us denote by x(μ) the minimizer of $\psi(x,\mu)$, and let (for μ > 0)

$$\lambda_i(\mu) = \frac{\mu}{c_i(x(\mu))}, \; i=1,2,\ldots m \tag{8}$$

Then $\nabla_x L(x(\mu), x(\mu)) = \nabla_x \psi = 0.$ Also, $c(x) > 0, \lambda > 0.$ > 0. So, all KKT conditions hold except for complementarity. In place of complementarity we have

$$c_i(x(\mu))\lambda_i(\mu) = \mu > 0, \qquad i = 1, 2, \ldots, m.$$

We are therefore on the center path defined artifcially by (8),

$$C_{pd} = \{(x(\mu))\lambda(\mu)s(\mu)) | \mu > 0\}, \tag{9}$$

Where s = c(x) are the slack variables. For the primal-dual formulation we can therefore write the above as

$$\nabla f(x) - A(x)^T \lambda = 0, \tag{10a}$$

$$c(x) - s = 0, \tag{10b}$$

$$\Lambda Se = \mu e, \tag{10c}$$

$$\lambda > 0, s > 0, \tag{10d}$$

Where, as in Section 8.2, Λ and S are diagonal matrices with entries λ_i and s_i, respectively, and $e(1,1,\ldots,1)^T$.

The setup is therefore that of primal-dual interior point methods. A modified Newton iteration for the equalities in (10) reads

$$\begin{pmatrix} \nabla_x^2 L(x,\lambda) & -A^T(x) & 0 \\ A(x) & 0 & -I \\ 0 & S & \Lambda \end{pmatrix} \begin{pmatrix} \delta x \\ \delta\lambda \\ \delta s \end{pmatrix} = \begin{pmatrix} -\nabla f + A^T \lambda \\ s - c \\ \mu e - \Lambda S e + r_{\lambda,s} \end{pmatrix}. \tag{11}$$

The modification term $r_{\lambda,s}$ (which does not come out of Newton's method-ology at all!) turns out to be crucial for both theory and practice.

Upon solving (11) we update

$$x \leftarrow x + \alpha\delta x, \lambda \leftarrow \lambda + \alpha\delta\lambda, s \leftarrow s + \alpha\delta s,$$

Choosing α so that the inequalities (10d) are obeyed.

3 Augmented Lagrangian Method

Consider the problem with only equality constraints, (1). The basic difficulty with the quadratic penalty method has been that elusive limit of dividing 0 by 0. Let us therefore consider instead adding the same penalty term to the Lagrangian, rather than to the objective function.

Thus, define the augmented Lagrangian

$$L_A(x,\lambda,\mu) = L(x) + \frac{1}{2\mu}\sum_{i=1}^{m} c_i^2(x),$$

$$= f(x) - \sum_{i=1}^{m} \lambda_i c_i(x) + \frac{1}{2\mu}\sum_{i=1}^{m} c_i^2(x). \tag{12}$$

The KKT conditions require, to recall, that $\nabla_x L(x^*,\lambda^*) = 0$ and $c(x^*) = 0,$ c(x.) = 0, so at the optimum the augmented Lagrangian coincides with the Lagrangian, and μ no longer need be small.

In fact, at some non-critical point,

$$\nabla_x L_A(x,\lambda;\mu) = \nabla f(x) - \sum_{i=1}^{m}\left[\lambda_i - \frac{c_i(x)}{\mu}\right]\nabla c_i(x),$$

hence we expect near a critical point that

$$\lambda_i - \frac{c_i(x)}{\mu} \approx \lambda_i^*,\ i=1,\ldots..,m. \tag{13}$$

We can choose some μ > 0 not very small. The minimization of the augmented Lagrangian (12) therefore yields a stabilization, replacing $\nabla^2 f$ by $\nabla^2 f + \frac{1}{\mu}A^T A.$ Thus, the Hessian matrix (wrto x) of the augmented

Lagrangian, $\nabla_x^2 L_A$, is s.p.d. provided that the reduced Hessian matrix of the

Lagrangian, $\mathrm{Z}^{\mathrm{T}}\left(\nabla_x^2 \mathrm{L}\right)\mathrm{Z},$ is s.p.d. It can further be shown that for μ small enough the minimization of (10.12) with respect to *x* at $\lambda = \lambda^*$ yields x^*.

Moreover, the formula (13) suggests a way to update λ_i in a penalty-like sequence of iterates:

$$\lambda_i^{k+1} = \lambda_i^k - \frac{c_i(x_k^*)}{\mu_k}.$$

In fact, we should update also μ while updating λ. This then leads to the following algorithmic framework:

Given $\mu_0 > 0$, x_0, and a final tolerance *tol*,

For $k = 0,1,2,\ldots.$

Starting with x_k minimize $\|\nabla L_A(x,\lambda_k;\mu_k)\|$, , terminating when

$$\|\nabla L_A(x,\lambda_k;\mu_k)\| \le \tau_k$$

Call the result x_k^*..

If final convergence test holds (e.g. $\tau_k \le tol$) exit

Else

$$\lambda_{k+1} = \lambda_k - \mu_k^{-1} c(x_k^*)$$

Choose $\mu_{k+1} \in (0, \mu_k)$

Choose a new starting point x_{k+1}, e.g. $x_{k+1} = x_k^*$.

This allows a gentler decrease of μ: both primal and dual variables par-ticipate in the iteration. The constraints satisfy

$$\frac{c_i(x_k^*)}{\mu_k} = \lambda_i^k - \lambda_i^{k+1} \to 0, \quad 1 \le i \le m,$$

Clear improvement over the expression (3) relevant for the quadratic penalty method.

Example 3

Let us repeat the experiments of Example 1 using the Augmented Lagrangian method. We use the same parameters and starting points, with $\lambda_0 = 0$.

For the quadratic 4-variable objective function we obtain convergence after a total of 29 Newton iterations. No damping was needed. The resulting penalty parameter sequence is

μ = 1,.1,.01,...,1.e-5,

and the corresponding Lagrange multiplier estimates are

$$\lambda = 0, -2.09, -12.78, -32.85, -40.59, -40.94.$$

For the non-quadratic objective function we obtain con-vergence after a total of 28 iterations. The penalty parameter sequence is

μ = 1,.1,.01,.001,1.e-4,

and the corresponding Lagrange multiplier estimates are

$$\lambda = 0, 9.7\mathrm{e}-7, -2.63, -3.33, -3.35.$$

The smallest values of μ required here are much larger than in Example 1, and no difficulty is encountered in the path to convergence for the augmented Lagrangian method: the advantage over the penalty method of Example 1 is more than the iteration counts alone indicate.

It is possible to extend the augmented Lagrangian method directly for inequality constraints [26]. But instead we can use slack variables. Thus, for a given constraint

$$c_i(x) \ge 0, \quad i \in I,$$

we write

$$c_i(x) - s_i = 0, \quad \mathrm{s}_i \geq 0.$$

For the general problem with equality con-straints plus nonnegativity constraints

$$\min_{x,s} \; f(x), \tag{14a}$$

$$\mathrm{c}_i(x) = 0, i \in \varepsilon, \tag{14b}$$

$$\mathrm{c}_i(x) - s_i = 0, i \in I, \tag{14c}$$

$$s \geq 0. \tag{14d}$$

For the latter problem we can utilize a mix of the augmented Lagrangian method applied for the equality constraints and the gradient projection method, as described in Section 9.3, applied for the nonnegativity constraints.

This is the approach taken by the highly successful general-purpose code LANCELOT by Conn, Gould and Toint [11]. In the algorithmic framework presented earlier we now have the sub problem

$$\min_{x,s} L_A(x, s, \lambda; \mu) \tag{15}$$

$$s.t. \; \mathrm{s} \geq 0,$$

Where λ and μ are held fixed when (15) is solved.

REFERENCES

1. *Bertsekas, D.P., (1976), "Multiplier Methods: A Survey", Automatica, Vol. 12, pp. 133-145.*
2. *Bertsekas, D.P., (1980a), "Enlarging the Region of Convergence of Newton's Method for Constrained Optimization", LIDS Report R-985, M.I.T., Cambridge, Mass. (to appear in J.O.T.A.).*
3. *Bertsekas, D.P., (1980b), "Variable Metric Methods for Constrained Optimization Based on Differentiable Exact Penalty Functions", Proc. of Eigtheenth Allerton Conference on Communication, Control and Computing, Allerton Park, Ill., pp. 584-593.*
4. *Bertsekas, D.P., (1982), Constrained Optimization and Lagrange Multiplier Methods, Academic Press, N.Y.*
5. *Boggs, P.T., and Tolle, J.W., (1980), "Augmented Lagrangians which are Quadratic in the Multiplier", J.O.T.A., Vol. 31, pp. 17-26.*
6. *Boggs, P.T., and Tolle, J.W., (1981), "Merit Functions for Nonlinear Programming Problems", Operations Research and Systems Analysis Report, Univ. of North Carolina, Chapel Hill.*
7. *Chamberlain, R.M., Lemarechal, C., Pedersen, H.C., and Powell, M.J.D., (1979), "The Watchdog Technique for Forcing Convergence in Algorithms for Constrained Optimization", Presented at the Tenth International Symposium on Mathematical Programming, Montreal.*
8. *DiPillo, G and Grippo, L., (1979), "A New Class of Augmented Lagrangians in Non-linear Programming", SIAM J. on Control and Optimization, Vol. 17, pp. 618-628.*

9. *DiPillo, G., Grippo, L., and Lampariello, F., (1979), "A Method for Solving Equality Constrained Optimization Problems by Unconstrained Minimization", Proc.9th IFIP Conference on Optimization Techniques, Warsaw, Poland.*

10. *Dixon, L.C.W., (1980), "On the Convergence Properties of Variable Metric Recursive Quadratic Programming Methods", Numerical Optimization Centre Report No. 110, The Hatfield Polytechnic, Hatfield, England. Fletcher, R., (1970), "A Class of Methods for Nonlinear Programming with Termination and Convergence Properties", in Integer and Nonlinear Programming (J.Abadie, ed.),North-Holland, Amsterdam.*

11. *Han, S.-P., (1977), "A Globally Convergent Method for Nonlinear Programming",.J.O.T.A., Vol. 22, pp. 297-309.*

12. *Han, S.-P., and Mangasarian, O.L., (1981), "A Dual Differentiable Exact Penalty Function", Computer Sciences Tech. Report #434, University of Wisconsin.*

13. *Maratos, N., (1978), "Exact Penalty Function Algorithms for Finite Dimensional and Control Optimization Problems", Ph.D. Thesis, Imperial College of Science and Technology, University of London.*

14. *Mayne, D.Q., and Polak, E., (1978), "A Super linearly Convergent Algorithm for Constrained Optimization Problems", Research Report 78-52, Department of Computing and Control, Imperial College of Science and Technology, University of London.*

15. *Powell, M.J.D., (1978),"Algorithms for Nonlinear Constraints that Use Lagrangian Functions", Math. Programming, Vol. 14, pp. 224-248.*

16. *Pschenichny, B.N., (1970), "Algorithms for the General Problem of Mathematical Programming", Kibernetica, pp. 120-125 (Translated in Cybernetics, 1974).*

17. *Pschenichny, B.N., and Danilin, Y.M., (1975),Numerical Methods in extremal Problems, M.I.R. Publishers, Moscow (English Translation 1978).*

18. *Rockafellar, R.T,, (1976), "Solving a Nonlinear Programming Problem by Way of a Dual Problem", Symposia Mathematica, Vol. XXVII, pp. 135-160.*

19. *Qashqaei, Amir, and Ramin Ghasemi Asl. "Numerical modeling and simulation of copper oxide nanofluids used in compact heat exchangers." International Journal of Mechanical Engineering, 4 (2), 1 8 (2015).*

20. *Lakshmi, B., and Z. Abdulaziz al Suhaibani. "A Glimpse at the System of Slavery." International Journal of Humanities and Social Sciences 5.1 (2016): 211.*

21. *Clementking, A., et al. "Neuron Optimization for Cognitive Computing Process Augmentation." International Journal of Computer Science and Engineering (IJCSE) ISSN 2278-9960 Vol. 2, Issue 3, July 2013, 5-12*